PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-146760
(43)Date of publication of application : 06.06.1995

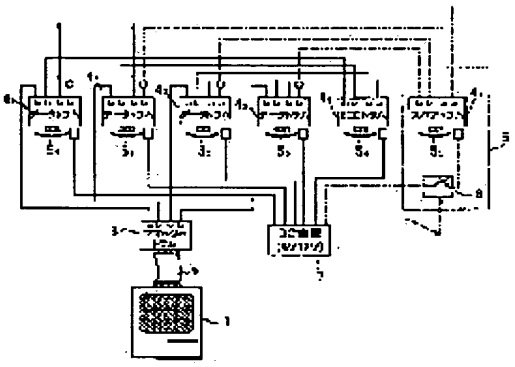
(51)Int.Cl. G06F 3/06
G06F 3/06

(21)Application number : 05-314487
(22)Date of filing : 19.11.1993

(71)Applicant : MUTOH IND LTD
(72)Inventor : UMITAKE TAKAFUMI

(54) DISK ARRAY DEVICE

(57)Abstract:
PURPOSE: To provide a disk array device by which a data recovery processing is automatically performed without interposing an operator.
CONSTITUTION: A spare disk device 55 is preliminarily provided in addition to disk devices 50 to 53 dispersedly storing data and a disk device 54 storing parity information ECC. If the data destruction in either one of the disk devices 50 to 53 is detected, the data of the disk device where data is destroyed is automatically restored from the data of other disk device based on the error detection result and the data is written in the spare disk 55.



LEGAL STATUS

[Date of request for examination]	27.11.1998
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3009987
[Date of registration]	03.12.1999
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号

特開平7－146760

(43)公開日　平成７年(1995)６月６日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	５４０			
	３０５　Ｚ			

審査請求　未請求　請求項の数３　ＦＤ　（全　５　頁）

(21)出願番号　特願平5－314487

(22)出願日　平成５年(1993)11月19日

(71)出願人　000238566
武藤工業株式会社
東京都世田谷区池尻３丁目１番３号

(72)発明者　海嶽　尚文
東京都世田谷区池尻３丁目１番３号　武藤工業株式会社内

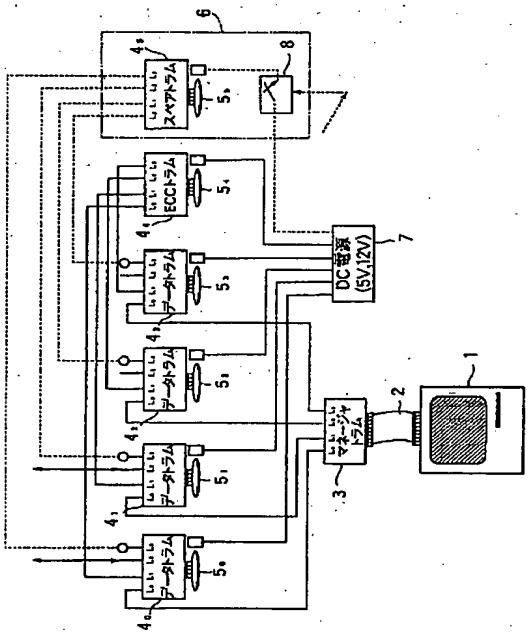
(74)代理人　弁理士　伊丹　勝

(54)【発明の名称】　ディスクアレイ装置

(57)【要約】

【目的】　オペレータの介在なしに自動的にデータ復帰処理がなされるようにしたディスクアレイ装置を提供する。

【構成】　データを分散記憶するディスク装置５０～５３及びパリティ情報ＥＣＣを記憶するディスク装置５４の他に、スペアディスク装置５５が予め設けられている。ディスク装置５０～５３のいずれかでのデータ破壊が検出されたら、そのエラー検出結果に基づいてデータが破壊がされたディスク装置のデータが他のディスク装置のデータから自動復元されて、スペアディスク装置５５に書き込まれる。



1

【特許請求の範囲】

【請求項1】　データを分散記憶する複数台のディスク装置を配置して構成されるディスクアレイ装置において、

前記複数台のディスク装置とは別個に設けられた予備ディスク装置と、

前記複数台のディスク装置でのデータ破壊を検出するエラー検出手段と、

このエラー検出手段の出力に基づいてデータ破壊が検出されたディスク装置のデータを他のディスク装置のデータから復元して前記予備ディスク装置に格納するデータ復元処理手段とを有することを特徴とするディスクアレイ装置。

【請求項2】　前記データ復元処理手段は、通常動作の合間にデータの復元処理を実行することを特徴とする請求項1に記載のディスクアレイ装置。

【請求項3】　前記データ復元処理手段は、データ復元処理時間間隔又はデータ復元処理1回の復元情報量の大きさ単位の少なくとも一方を可変として、通常動作の合間にデータの復元処理を実行する機能を有することを特徴とする請求項1に記載のディスクアレイ装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、データを分散記憶する複数台のディスク装置を配置して構成されるディスクアレイ装置に関する。

【0002】

【従来の技術】大量のデータを複数台の小型ディスク装置に分散して記憶することにより危険分散を図るディスクアレイ装置として、ＲＡＩＤ（Redundant Arrays of Inexpensive Disks）が知られている。ＲＡＩＤには、1から5までのレベルがある。2～4のレベルでは、複数のディスク装置がデータを分散記憶するものと、エラーチェック用のパリティ情報を記憶するものとに用途分けされている。レベル5では全てのディスク装置にデータと共にパリティ情報も分散して記憶される。このようなＲＡＩＤ装置の中のいずれかのディスクでデータが破壊されたとき、従来は故障ディスクに代わって新しいディスクを装填し、データ復帰ボタン等により復帰プログラムをスタートさせ、破壊されたデータ分を新しいディスクに復帰させる方式が採られている。

【0003】

【発明が解決しようとする課題】従来のＲＡＩＤでのデータ復帰方式は、オペレータにとって新しいディスクの装填とデータ復帰という操作を必要とし、操作ミスが生じ易い。また、このデータ復帰処理の間、ディスク装置のアクセスは中止しなければならない。本発明は、オペレータの介在なしに自動的にデータ復帰処理がなされるようにしたディスクアレイ装置を提供することを目的としている。

2

【0004】

【課題を解決するための手段】本発明は、データを分散記憶する複数台のディスク装置を配置して構成されるディスクアレイ装置において、前記複数台のディスク装置でのデータ破壊を検出するエラー検出手段と、このエラー検出手段の出力に基づいてデータ破壊が検出されたディスク装置のデータを他のディスク装置のデータから自動復元して格納するための予備のディスク装置とを有することを特徴としている。本発明において好ましくは、データ復元処理手段は、通常動作の合間にデータの復元処理を実行する。この場合、効率的な復元処理を行うためには、データ復元処理の時間間隔を可変とする機能、又は1回のデータ復元処理で復元すべき情報量の大きさ単位を可変とする機能の少なくとも一方をオプション機能としてシステムに組み込むことが有効である。

【0005】

【作用】本発明によれば、ディスクアレイ装置には予め予備のディスク装置が設けられていて、あるディスクにデータエラーがあった場合には、自動的にその故障ディスクのデータが他のディスクのデータによって復元されて予備のディスク装置に格納される。従って、大切なデータがオペレータの操作ミスで復帰できなくなるといった事態が防止できる。特に本発明を、並列データ転送可能な通信プロセッサ（以下、トラムという）をツリー状に配置して複数のディスク装置の並列処理を可能としたシステムに適用すれば、トラム間通信によってデータエラー検出後直ちに、通常動作と並列にデータ復元処理ができる。また、ディスク装置とホストコンピュータとのデータ授受の合間を利用してデータ復元処理を行うことにより、通常動作に何等影響を与えることなくデータの復元が可能になる。

【0006】

【実施例】以下、図面を参照して、本発明の実施例を説明する。図1は、本発明の一実施例に係るディスクアレイ装置のブロック構成を示す。この装置は、5台のディスク装置50～54を有する。これらのディスク装置50～54のうち4台のディスク装置50～53がデータを分散記憶するものであり、残り1台のディスク装置54はパリティ情報ＥＣＣを記憶するものである。これらのディスク装置50～54に1対1に対応してアクセス手段としてのデータトラム40～43及びＥＣＣトラム44が設けられている。これらのトラム40～44は例えば、並列データ転送が可能な4つのシリアルポートを持つ通信プロセッサであり、好適にはインモス社のトランピュータ（商標）等が使用される。またこれらのトラム40～44にはそれぞれＳＣＳＩポート（パラレル）が設けられており、このＳＣＳＩポートを介してそれぞれディスク装置50～54に接続されている。

【0007】ホストコンピュータ1をアクセスするデータ転送手段として、マネージャトラム3が設けられてい

る。このマネージャトラム３もデータトラム４０～４３及びＥＣＣトラム４４と同様の４つのシリアルポートを持つ通信プロセッサである。マネージャトラム３のパラレルポートは、ＳＣＳＩインタフェース２を介してホストコンピュータ１と接続されている。マネージャトラム３の４つのシリアルポートは、データトラム４０～４３の各１つのシリアルポートと接続されている。ＥＣＣトラム４４の４つのシリアルポートは、データトラム４０～４３の各１つのシリアルポートと接続されている。

【０００８】以上の基本構成の他、ＲＡＩＤ装置内には、ディスク装置に故障があった場合の予備装置６として、スペアディスク装置５５及びスペアトラム４５が設けられている。スペアトラム４５も、データトラム４０～４３及びＥＣＣトラム４４と同様の４つのシリアルポートを持つ通信プロセッサであり、その４つのシリアルポートは破線で示すようにそれぞれデータトラム４０～４３の各１つのシリアルポートに接続されている。この予備装置６の電源スイッチ８は、ＲＡＩＤ装置が正常動作している限りオフであり、通常直流電源７の出力は基本構成部のみに供給される。ディスク装置に故障があった場合には、マネージャトラム３の制御により、又はエラーメッセージに従ったオペレータの手動操作より電源スイッチ８がオンされる。

【０００９】この実施例のＲＡＩＤ装置において、記憶すべきデータは４つのディスク装置５０～５３にビット単位又はセクタ単位に分割されて書き込まれる。残りのディスク装置５４には、これらの分割書込みされたデータのパリティ情報ＥＣＣが書き込まれる。パリティ情報ＥＣＣは例えば、マネージャトラム３で計算される。このデータの書込み／読出しの詳細は省略する。

【００１０】この実施例のＲＡＩＤ装置において、ディスク装置５０～５３のいずれかでのデータ破壊はＳＣＳＩインタフエース信号により検出することができる。データ破壊が検出されたら、残りのディスク装置のデータに基づいて破壊されたデータの内容を自動的に復元してこれをスペアディスク装置５５に格納するという処理がなされる。この処理は例えば、マネージャトラム３のデータ復元プログラムによってなされる。あるいはスペアトラム４５でこの処理を行うこともできる。

【００１１】データ復元処理の流れは、簡単に示せば図２の通りである。ＳＣＳＩインタフェース信号からディスク装置のエラー検出がなされる。エラーが検出されたら、この検出結果によりエラーステータスをオンにする（Ｓ１）。そして破壊されたディスク装置の電源スイッチをオフ（Ｓ２）、スペアディスク装置５５の電源スイッチ８をオン（Ｓ３）して、データ復元処理を行う（Ｓ４）。破壊されたデータが全て復旧されたら、エラーステータスをオフにし（Ｓ５）、以後通常動作に戻る。

【００１２】破壊されたディスクの内容は前述のように他のディスクから逆計算されるが、そのデータ復元処理

のアルゴリズムは次の通りである。４つのディスク装置５０～５３のデータをそれぞれＡ，Ｂ，Ｃ，Ｄで表すと、パリティ情報ＥＣＣは、次の数１で示される。なお以下の式で、”＋”は排他的論理和（ＸＯＲ）を表す。

【００１３】

【数１】Ａ＋Ｂ＋Ｃ＋Ｄ＝ＥＣＣ

【００１４】データＢが破壊されたと仮定して、これを残りのデータＡ，Ｃ，Ｄ，ＥＣＣから求める場合を説明すれば、数１の両辺にＥＣＣ＋ＢをＸＯＲすると、

【００１５】

【数２】

Ａ＋Ｂ＋Ｃ＋Ｄ＋ＥＣＣ＋Ｂ＝ＥＣＣ＋ＥＣＣ＋Ｂ

【００１６】となる。数２は、符号理論から、Ｂ＋Ｂ＝０，ＥＣＣ＋ＥＣＣ＝０であるから、結局次の数３と等価になる。

【００１７】

【数３】Ａ＋Ｃ＋Ｄ＋ＥＣＣ＝Ｂ

【００１８】従って数３に基づいて、破壊されたデータＢが逆計算されて求められることになる。

【００１９】この実施例において、破壊されたデータの復元処理は、例えば図３に示すようにホストコンピュータ１とのデータ授受の合間を利用して１セクタデータずつ行うことができる。この動作において、破壊されたデータの全てが復旧されたとき、エラーステータスをオフにしてエラーディスクのない通常動作を行う。これにより通常のデータ読み書き動作に何等影響を与えることなく、データ復元処理を実行することができる。また、各トラム３，４０～４５に並列動作可能な通信プロセッサを用いれば、データ修復をトランザクションと並列に実行することが可能である。

【００２０】通常動作の合間を利用したデータ復元処理の場合、復元処理中にもしトランザクション要求が発生したとすると、トランザクションが優先される。従ってデータ復元処理中に頻繁にトランザクションが行われると、復元処理完了までに多くの時間がかかる。この問題を解決するには、次の二つの機能のいずれかをオプション機能としてシステムに組み込むことが有効である。

①データ復元処理の時間間隔を可変とする機能

②１回のデータ復元処理で復元すべき情報量の大きさ単位を可変とする機能

【００２１】①の方法は、復元処理の時間間隔Ｔを、図４（ａ）に例示したＴ１あるいはＴ２のように可変とするものであり、②の方法は、１回の復元処理で復元すべき情報量の大きさの単位Ｒを、図４（ｂ）に例示したＲ１あるいはＲ２のように可変とする。これらの具体的な機能実現方法としては、例えばディップスイッチで選択してその内容をシステムソフトが読むという方法を用いることができる。

【００２２】本発明は上記実施例に限られない。実施例ではパリティ情報ＥＣＣをデータと別に固定のディスク

5

に記憶するレベル２～４のＲＡＩＤを説明したが、ＥＣＣをデータと同様に分散記憶するレベル５にも同様に本発明を適用することができる。

【００２３】

【発明の効果】以上述べたように本発明によれば、予備のディスク装置を備えることで、データ破壊があった場合にオペレータの介在なしに自動的にデータ復帰処理がなされるようにして、ディスクアレイ装置の操作ミスを防止することができる。

【図面の簡単な説明】

【図１】　本発明の一実施例に係るディスクアレイ装置を示す。

6

【図２】　同実施例のデータ復元処理の流れを示す。

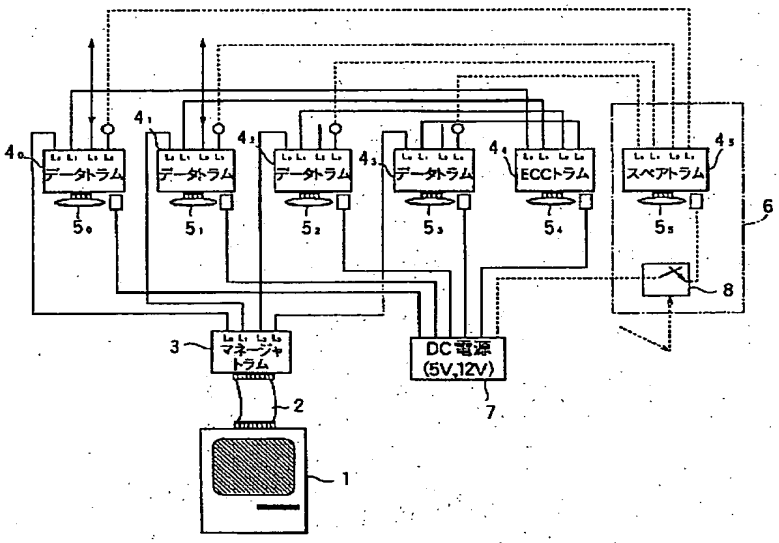
【図３】　同実施例のデータ復元の並列処理の様子を示す。

【図４】　他の実施例のデータ復元の並列処理の様子を示す。

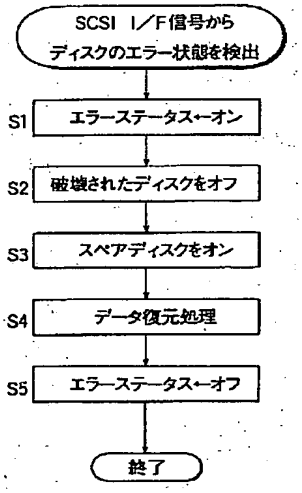
【符号の説明】

１…ホストコンピュータ、２…ＳＣＳＩインタフェース、３…マネージャトラム、４０～４３…データトラム、４４…ＥＣＣトラム、５０～５４…ディスク装置、４５…スペアトラム、５５…スペアディスク装置、７…直流電源、８…電源スイッチ。

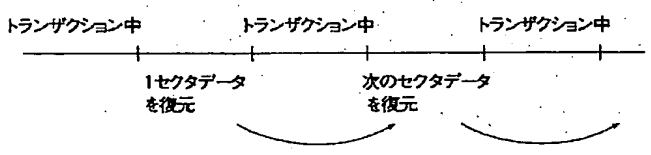
【図１】



【図２】

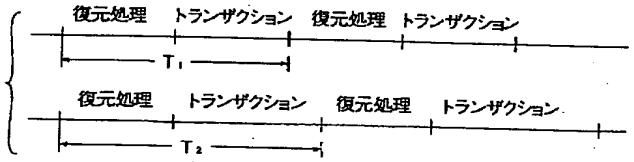


【図３】



【図４】

(a) 復元処理時間間隔可変



(b) 復元処理単位情報量可変